LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Cognitive and Sociopsychological Individual Differences, Experience, and Naturalistic Second Language Speech Learning: A Longitudinal Study

Hui Sun [a], Kazuya Saito [b], and Jean-Marc Dewaele [c]

[a]Cardiff University [b]University College London [c]Birkbeck, University of London

Abstract: This study longitudinally examined the effects of cognitive and sociopsychological individual differences (aptitude, motivation, personality) and the quantity and quality of second language (L2) experience on L2 speech gains in naturalistic settings. We elicited L2 spontaneous speech from 50 Chinese learners of English at the beginning and the end of their first 4 months of study abroad. Then, we linked the participants' gains in comprehensibility (ease of understanding) and accentedness (linguistic nativelikeness) to their individual difference and experience profiles. The participants' gains in comprehensibility were associated mainly with the amount of their interaction with fluent English speakers during immersion and secondarily with certain

CRediT author statement – **Hui Sun**: conceptualization; methodology; funding acquisition; investigation; formal analysis; writing – original draft preparation; writing – review & editing. **Kazuya Saito**: conceptualization; methodology; supervision; writing – review & editing. **Jean-Marc Dewaele**: conceptualization; methodology; supervision; writing – review & editing.

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

The project was funded by a *Language Learning* dissertation grant. The results reported here are based on the first author's doctoral dissertation submitted to Birkbeck, University of London in 2019. We would like to thank anonymous *Language Learning* reviewers for their insightful comments on earlier versions of the manuscript and Editors Emma Marsden and Pavel Trofimovich for their support throughout the review, revision, and production.

Correspondence concerning this article should be addressed to Hui Sun, Cardiff University, John Percival Building, Colum Drive, Cardiff, CF10 3EU, United Kingdom.
Email: sunh21@cardiff.ac.uk

The handling editor for this manuscript was Emma Marsden.

DOI: 10.1111/lang.12561

cognitive (grammatical inferencing) and sociopsychological (extraversion) individual differences. Furthermore, the amount of interactive L2 use mediated the effect of sociopsychological individual differences (extraversion and potentially ideal L2 self). In contrast, gains in accentedness tended to be less subject to experience effects but could be affected by certain pronunciation-related cognitive individual differences (phonemic coding).

Keywords second language speech; aptitude; motivation; personality; experience; study abroad; pronunciation

Introduction

Over the past decade, two global dimensions of second language (L2) speech—comprehensibility and accentedness, typically operationalized through naïve listeners' intuitive judgements—have been increasingly studied in L2 speech research (e.g., Derwing & Munro, 2013; Huensch & Nagle, 2021; Saito & Plonsky, 2019; Trofimovich et al., 2022). Comprehensibility refers to how much effort it takes to understand a L2 learner's speech, whereas accentedness is conceptualized as how different a L2 learner's speech sounds compared to the local variety. These two constructs have been found to only partially overlap, as L2 speech with a heavy foreign accent is not necessarily difficult to comprehend (Munro & Derwing, 2020).

Research on L2 speech learning has shown that high-quality experience, that is, sufficient, naturalistic, and interactive exposure to the target language, is one of the most crucial variables in the acquisition of advanced L2 speech proficiency (e.g., Moyer, 2011; Saito, 2015). In addition to learner-external variables, learner-internal variables—cognitive and sociopsychological individual differences (IDs), especially aptitude and motivation—have also been related to L2 speech comprehensibility and accentedness (for aptitude, see Granena & Long, 2013; Saito et al., 2019; for motivation, see Nagle, 2018; Saito et al., 2019). Researchers in L2 acquisition have called for a multidimensional examination of the role of IDs (Moyer, 2014; Ortega, 2009). For example, Moyer (2014) argued in her review of recent qualitative studies that exceptional nativelike accent attainment is the result of a constellation of cognitive, sociopsychological, and experiential variables. To our knowledge, however, no quantitative, longitudinal study has been done to investigate how the three types of ID variables can uniquely explain the development of comprehensible and nativelike L2 speech in naturalistic settings.

In this empirical study, we therefore longitudinally examined the L2 English comprehensibility and accentedness of 50 Chinese international students

at a UK university over the first 4 months of immersion, exploring the influence of these students' L2 experience (quantity and quality) and cognitive and sociopsychological IDs (aptitude, motivation, and personality) on L2 speech gains (i.e., changes in L2 speech performance over time).

Background Literature

Comprehensibility and Accentedness of L2 Speech

The distinction between comprehensibility and accentedness is crucial for various reasons (for a meta-analysis, see Saito, 2021). Whereas many L2 learners tend to take nativelikeness as an idealized goal, empirical studies have shown that few learners can attain nativelike pronunciation proficiency (Moyer, 2013). This has led to the importance of setting a realistic goal such as attaining comprehensible speech that interlocutors can successfully and easily understand (Munro & Derwing, 2020). While L2 learners with a strong orientation toward nativelikeness should not be discouraged from pursuing their own goals, students and teachers alike are strongly recommended to be aware of the facts (a) that many successful L2 learners' speech is accented but highly comprehensible and intelligible (Munro & Derwing, 2020) and (b) that it is efficient and effective to prioritize certain aspects of language with a view of attaining comprehensibility rather than nativelikeness (Trofimovich et al., 2022).

Indeed, accentedness seems to require more L2 use—and possibly extra instruction or training—for learners to improve compared with comprehensibility. For example, in a longitudinal study of L2 speech development over 7 years of immersion, Derwing and Munro (2013) found that L2 learners' accentedness developed moderately with a leveling-off at the end of the second year. In contrast, the learners' comprehensibility demonstrated significant, robust, and extensive progress throughout the 7 years (i.e., beyond the initial immersion phase) among the learners who had ample opportunities to use the language, especially with first language (L1) English interlocutors. To further this line of research, our study used fine-grained measures of both quantity (weekly hours of L2 use) and quality (L2 use in different modalities, with different types of interlocutors) of L2 experience. We expected to replicate previous findings that good-quality L2 experience (e.g., interactive L2 use with L1 interlocutors) would explain variance in comprehensibility to a higher degree than in accentedness.

A Multidimensional Model for L2 Speech Learning

The literature on experience effects on L2 speech learning (e.g., Saito, 2015) has suggested that experiential variables alone cannot fully explain the rate

and outcome of L2 speech acquisition. This in turn concurs with the view that L2 learning is a multidimensional phenomenon, affected by a range of learner-internal characteristics (Dörnyei, 2005). Moyer (2014, p. 435) emphasized the importance of "taking deliberate account of various social, cognitive, and psychological data in order to solidify a predictive model" for the acquisition of more advanced L2 speech proficiency. Moyer's multidimensional model consists of three crucial components affecting successful L2 speech learning: experiential variables (e.g., L2 use across multiple domains especially social interaction), cognitive variables (e.g., aptitude for mimicking or discriminating accent), and sociopsychological variables (e.g., motivation for assimilating linguistically and culturally, outgoing/extraverted personality).

Although Moyer (2014) proposed a model with a focus on L2 accentedness (or phonology), cognitive and sociopsychological variables could also explain comprehensibility-related domains such as lexicogrammar. For example, learners rely on their ability to identify patterns to learn morphosyntactic rules from input and rely on their associative memory to acquire new vocabulary. Language-related goals may help learners seek L2 input and attend to the input with a higher level of awareness and thus enhance input processing, which in turn may contribute to the acquisition of lexicogrammar. In fact, the effects of these cognitive and sociopsychological variables have been found in L2 comprehensibility development (e.g., Saito et al., 2019). Therefore, to fully understand the learning of L2 comprehensibility and accentedness, one needs to consider experiential, cognitive, and sociopsychological variables in a symbiotic manner.

However, few studies have included all three dimensions for investigating their relative contributions to L2 speech learning. Our study took a first step in applying the multidimensional approach by focusing on actual L2 use and a cluster of cognitive and sociopsychological ID constructs that are (most) extensively researched in L2 acquisition: (a) aptitude, (b) motivation, and (c) personality. These ID constructs were also found to be predictors of exceptional outcomes in L2 phonology in Moyer's model (2014). In what follows, we have provided a review for each of the constructs. It is worth noting that the ID variables tend to affect accentedness and comprehensibility in different ways, arguably due to the differences in the nature of accentedness and comprehensibility.

Aptitude and L2 Speech Learning

Language learning aptitude generally refers to a set of cognitive and perceptual abilities that help L2 learners learn a language effectively and efficiently

in various contexts. In line with his well-known aptitude test battery the Modern Language Aptitude Test, Carroll originally conceptualized aptitude as four different components: phonemic coding, grammatical sensitivity, inductive language learning, and associative memory (Carroll & Sapon, 1959). More recently, researchers have begun to expand the scope of aptitude framework by including a set of cognitive abilities relevant to both explicit and implicit language learning. Meara's (2005) computer-based LLAMA test battery (used in our study), for example, features three different aptitude constructs—phonemic coding, grammatical inferencing, and associative memory—to measure explicit learning aptitude (with intention) and one aptitude construct—novel sound recognition—to measure potentially incidental learning aptitude (without intention; but see Suzuki's, 2021, critiques of the distinction between intentional vs. incidental learning abilities in LLAMA).

Whereas researchers have continued to debate its internal and external validity (e.g., Bokander & Bylund, 2020; Granena, 2013), the LLAMA test has been widely used in L2 research. A number of empirical studies have shown that learners' explicit aptitude predicted the acquisition of various L2 lexicogrammar dimensions, especially the difficult, complex, and nonsalient features (S. Li, 2013), under foreign language learning conditions where language is taught and learned with the focus on form rather than on meaning (see S. Li, 2015, for a meta-analytic review). More recently, researchers have found evidence for different roles for explicit and incidental aptitude in adult L2 speech learning in classrooms (e.g., Granena, 2019; Saito, 2017) and in naturalistic settings (e.g., Granena & Long, 2013). Saito (2017) found L2 speech attainment to be related to explicit (but not implicit) aptitude among Japanese learners of English at various levels. Learners with better associative memory (linguistic information retaining) were more fluent in L2 English speech; those with stronger phonemic coding ability (phonetic input processing) were better at pronunciation; and those with stronger grammatical inferencing ability (morphosyntactic pattern identifying) tended to use more diverse vocabulary. However, in Granena's (2019) study of American learners of L2 Spanish at an intermediate level, incidental memory predicted speed fluency and lexical diversity to a small degree (but only among learners with high implicit learning ability), whereas explicit learning ability did not show significant effects. This difference in findings might be explained by the different L2 experience and levels of participants. According to Saito et al.'s (2019) longitudinal study, for example, a significant effect of incidental aptitude (novel sound recognition) was found on gains in L2 segmental accuracy at the later stage of learning (for similar results, see also Suzukida & Saito, 2022).

Compared with L2 speech learning in classrooms, few studies have investigated how L2 learners with different aptitude profiles improve their L2 speech in naturalistic settings. Granena and Long (2013) reported a significant correlation between composite aptitude scores (a sum score of four LLAMA subtests) and the ultimate L2 attainment in pronunciation and lexis among 18 Chinese learners at an advanced level. However, the cross-sectional design did not allow an examination of causal relationships, and Granena and Long could draw no conclusions for specific aptitude components. In this study, we took a longitudinal look at the effects of explicit and implicit aptitude on gains in L2 comprehensibility and accentedness over a short period of immersion. Based on the previous findings reviewed above, we predicted that L2 comprehensibility gains would be related to associative memory and grammatical inferencing ability, which could help refine L2 fluency and lexicogrammar (comprehensibility-related domains). However, we anticipated that L2 accentedness gains would be influenced by pronunciation-related aptitude, that is, phonemic coding ability (explicit aptitude) or sound recognition (implicit aptitude), depending on learners' L2 experience, meaning that implicit learning is more likely to happen with extensive high-quality L2 input.

Motivation and L2 Speech Learning

L2 motivation refers to a set of motives that drive a person to consistently make efforts to learn a target language. Motivation has been considered instrumental in L2 learning outcomes: Not only is it a prerequisite for functions of other variables, but it could also make up for their inadequacy. Dörnyei's (2005, 2009) L2 motivational self system (L2MSS) model, which is based on the possible selves theory (Markus & Nurius, 1986) and the self-discrepancy theory (Higgins, 1987), has dominated recent L2 research. According to the L2MSS model, learners can be motivated to invest more effort in L2 practice to decrease the discrepancy between their actual selves and two types of desired future selves as L2 users—ideal L2 self (one's future self with all attributes one aspires or expects to possess regarding the target language) and ought-to L2 self (one's future self that meets all expectations of others regarding the target language). In other words, when L2 learners are motivated to improve their L2 skills to fulfil their own interests, needs and goals (ideal L2 self) or others' expectations and hopes (ought-to L2 self), they intend to spend more time and energy practicing the target language (intended learning efforts), which could then lead to improvement in the target language.

Over the past 20 years, empirical studies have demonstrated the relationships between the two future self-guides measured by scales for the L2MSS

and intended learning efforts among L2 learners (see Boo et al., 2015, for a comprehensive review). Recent studies have further investigated whether the self-guides contribute to L2 learning achievement through intended learning efforts, but the findings have been rather inconsistent (see Al-Hoorie, 2018, for a meta-analysis). Among a few longitudinal studies, Nagle (2018) tracked 26 novice learners of Spanish at a US university over a year to assess the links between time-varying motivation and L2 speech development. Nagle found that intended learning efforts positively predicted L2 development in comprehensibility and accentedness but neither self-guide showed any predictive power, and thus he suggested that the self-guides may play a more important role in an immersion context (p. 212). Interestingly, Saito et al. (2018) reported different results by tracking 122 high school learners of English in Japan over one term—ideal L2 self positively (and L2 use marginally) predicted L2 comprehensibility gains, suggesting a direct effect of ideal L2 self on L2 speech learning. However, Saito et al. did not examine the mediation effect of L2 use. Thus, in this study, we took a first step toward examining the degree to which the two future self-guides predict L2 speech gains in naturalistic settings (with extensive and authentic L2 input) and the degree to which actual L2 use mediated the effect (see Hernández, 2010, for an indirect effect of integrative motivation on L2 speech gains via L2 interaction in a study-abroad setting).

It is noteworthy that Papi et al. (2019; see also Papi & Khajavy, 2021; Teimouri, 2017) recently refined the L2MSS model by distinguishing the regulatory focuses and standpoints of the two self-guides. According to the revised model, items in the ideal L2 self scale should have a promotion regulatory focus (approaching positive outcomes) and items in ought-to L2 self scale should have a clear prevention focus (avoiding negative outcomes); and both self-guides should have items representing the own standpoint and the other standpoint. Thus, the original two self-guides have been split into four scales: ideal L2 self/own (corresponding to the original ideal L2 self), ought-to L2 self/own, and ideal L2 self/other and ought-to L2 self/other (together corresponding to the original ought-to L2 self). A recent study used this model to examine the link between motivation and L2 speech proficiency among a small group of English foreign language learners (Kermad, 2018), and the results were consistent with Saito et al.'s (2018) results—ideal L2 self/own scores were positively associated with fluency and segmental accuracy. Although all items used in our study were from the original L2MSS model (we collected the data before Papi et al., 2019, had proposed the revised model), we decided to relabel the scales to make comparisons easier in future studies. As the ideal L2 self items used are promotion-focused with an own standpoint (e.g.,

"I imagine myself as someone who is able to speak English"), we relabeled them ideal L2 self/own. Similarly, because the ought-to L2 self items used are promotion-focused with an other standpoint (e.g., "I study English because close friends of mine think it is important"), we relabeled them ideal L2 self/other. No items measured ought-to L2 self under the revised framework.

We expected that the ideal L2 self/own scale would predict comprehensibility gains, replicating findings in classroom settings based on spontaneous speech elicited from learners at similar proficiency level (Kermad, 2018; Saito et al., 2018) and that L2 use would mediate the effect of motivation on L2 gains because (a) learners should have a more vivid image of their desired future selves (i.e., a strong ideal L2 self/own) being surrounded by native speakers during immersion (Q. Li, 2014), (b) a stronger ideal L2 self/own could motivate them to have more L2 interaction, and (c) L2 comprehensibility learning tends to benefit from L2 use especially at the initial stage of immersion (Derwing & Munro, 2013; Saito, 2015). Compared with comprehensibility, acquiring a nativelike accent should not be as important to most L2 learners in a communication-focused context, thus the relationship between motivation and accentedness gains might be weaker. The role of ideal L2 self/other could be less clear, with learners potentially influenced less by their closest families and friends who live far away.

Personality and L2 Speech Learning

Another crucial sociopsychological domain of IDs is personality, which is generally conceived as a person's typical and consistent patterns of behavior, thoughts, and feelings (Pervin & Cervone, 2010). The dominant model of personality is the Big Five, referring to five broad personality dimensions or factors—extraversion (vs. introversion), neuroticism (vs. emotional stability), openness to experience, agreeableness, and conscientiousness (e.g., Goldberg, 1981). Researchers have commonly measured the five-factor model using the NEO Five-Factor Inventory (Costa & McCrae, 1989). More recently, van der Zee and van Oudenhoven (2000) developed the Multicultural Personality Questionnaire to measure five personality traits that relate particularly to effectiveness and success in a multicultural situation—cultural empathy, open-mindedness, social initiative, emotional stability, and flexibility. This model overlaps with Big Five to some degree, with open-mindedness corresponding to openness to experience, social initiative to extraversion, and emotional stability to low neuroticism. Given that the L2 learning context of our study was a multicultural environment (i.e., London, UK), we chose this questionnaire to examine whether certain personality traits could contribute to the success of L2 learning during immersion.

Research has shown that personality traits have a small effect on L2 learning behavior and outcome (see Chen et al., 2022, for a meta-analytic review). For example, Oz˙ańska-Ponikwia and Dewaele (2012) found that openness to experience (or open-mindedness), which reflects the willingness to explore and to appreciate new experience, predicted the frequency of L2 use and self-perceived L2 proficiency among Polish immigrants in Ireland and the UK. Similarly, learners who tend to actively approach social interaction (social initiative/extraversion) are likely to have more interactive L2 use. There is also evidence that L2 learners who are more stable emotionally (emotional stability) tend to show more positive attitudes toward L2 community (MacIntyre & Charos, 1996) and feel less anxious about using the target language (Dewaele & Al-Saraj, 2015). These individuals are generally more confident and comfortable about expressing themselves in L2 interaction and more likely to achieve more advanced oral proficiency (Pyun et al., 2014). Given that L2 comprehensibility tends to benefit from L2 use, learners with the above personality profiles may achieve more comprehensible L2 speech. As for accent, a few studies have reported a strong correlation between empathy (the ability to perceive and feel others' emotional states) and accent/pronunciation (e.g., Ibrahim et al., 2008; Rota & Reiterer, 2009). Literature on empathy has suggested its crucial role in the authenticity of L2 pronunciation (Guiora et al., 1972). Guiora et al. suggested that nativelike pronunciation is developed gradually during the stage of ego-permeability, where a new identity needs to be formed in order to identify the subtle speech nuances and then to produce them, and that the ability to empathize and merge with others' emotional experience could promote this process.

Given that studies on the relationship between personality and L2 learning have been exclusively cross-sectional, our study was the first to investigate longitudinally whether and to what degree learners with different personality profiles differentially improve their L2 speech after a short period of naturalistic L2 learning. On the basis of previous findings, we expected to see effects of certain interaction-related personality traits (e.g., open-mindedness, social initiative, emotional stability) on L2 comprehensibility gains, likely via the mediation of L2 use. For L2 accentedness gains, we anticipated that cultural empathy would play a role.

Current Study

Our study aimed to explore the extent to which a multidimensional model of cognitive and sociopsychological ID (aptitude, motivation, personality) and L2 experiential variables (Moyer, 2014) predict the gains in L2 speech

comprehensibility and accentedness made by 50 Chinese learners of English over a 4-month immersion in the UK. In L2 acquisition research, there has been a lack of clarity on what exactly constitutes a longitudinal design (Ortega & Iberri-Shea, 2005). Following current practice in psychology, we defined the longitudinal nature of this study as involving "repeated measures of (at least twice) the same variables gathered from the same study participants over time" (Dormann & Guthier, 2019, p. 146). To this end, we tracked and compared participants' linguistic profiles at the outset and endpoints of the four months. We addressed two research questions in this study:

1. To what extent do L2 experiential, cognitive, and sociopsychological ID variables predict L2 speech gains in early immersion?
2. If sociopsychological ID variables show any effect on L2 speech gains, to what extent do L2 experiential variables mediate the relationship?

Regarding Research Question 1, we hypothesized that the experiential and ID variables would predict the participants' L2 speech gains in comprehensibility and accentedness to different degrees and in different ways. For example, both experiential and ID variables could explain variance in comprehensibility gains, whereas mainly ID variables could predict variance in accentedness gains. More specifically, learners who demonstrate greater comprehensibility gains are likely engaged in more interaction with fluent speakers (Derwing & Munro, 2013). Such regular L2 users tend to have a clearer image of ideal L2 self (own standpoint; Saito et al., 2018), initiate more social interaction, and be more culturally open-minded and emotionally stable (Ożańska-Ponikwia & Dewaele, 2012; Pyun et al., 2014). Gains in accentedness, however, would be less sensitive to L2 experience (Saito, 2015), and thus experiential variables would predict gains in accentedness to a lesser degree. Learners who are more culturally empathetic (Ibrahim et al., 2008), however, might improve more in accent. Besides, stronger aptitude related to fluency (associative memory) and lexicogrammar (grammatical inferencing) could benefit learners in comprehensibility gains, whereas accentedness gains could be tied to individual differences in participants' pronunciation-related (phonemic coding, novel sound recognition) aptitude (Granena & Long, 2013; Saito, 2017). For Research Question 2, as we suggested above, we expected the amount of interactive L2 use to mediate the effect of sociopsychological ID variables—measured by the ideal L2 self/own, social initiative, open-mindedness, and emotional stability scales—on gains in comprehensibility to some extent.

Method

Participants

We recruited 60 Chinese students (52 females, 8 males) within their first month of study-abroad at universities in London, UK. We contacted them via a platform where Chinese students exchange information with study-abroad peers before they arrive in the UK. Ten of them did not participate in the second session, which left 50 participants (44 females, 6 males). The participants were enrolled in tertiary-level courses in various subjects (e.g., economics, science, law, linguistics, arts); none of them took any English language courses during this 4-month research project. The participants were young adult learners ($M_{age} = 22.8$ years, *range* = 20–33) who had studied English as a foreign language for 10–16 years in China with very few opportunities to interact with L1 English speakers. None reported any extensive experience (i.e., over one month) in English-speaking countries prior to their arrival in the UK. Their self-reported IELTS scores varied widely from 5.5 to 8 out of 9 ($M = 7.12$, $SD = 0.48$), indicating that their overall L2 English proficiency ranged from independent (B1/B2) to proficient (C1) according to the Common European Framework of Reference for Languages.

We tested the participants twice individually in a quiet room at Birkbeck, University of London, at the outset and the end of the project. In the first session (T1), the participants completed the aptitude test, followed by an English-speaking test (picture narrative), a motivation questionnaire, and a personality questionnaire. After a 4-month interval, in a second session (T2), the participants took the speaking test (but with a different set of pictures) and answered questions about their L2 use over the T1–T2 period. The first session took approximately 60 minutes and the second session approximately 30 minutes. Figure 1 summarizes the timeline for participant testing.

L2 Speech Analyses

Material Preparation

Following practices found in the L2 speech literature (e.g., Derwing & Munro, 2013; Saito, 2015), we used picture narrative tasks to elicit participants' spontaneous speech. We adapted the materials (a four-frame cartoon story) from the Pre-Grade 1 Level of the EIKEN English Test (EIKEN Foundation of Japan, 2016). We selected the speech task to accommodate participants of varying English proficiency levels (Common European Framework of Reference for Languages levels B1–C1). With a storyline provided, the participants could focus more on language production, and the speech that they produced would be more likely comparable. In accordance with the testing procedure in EIKEN,

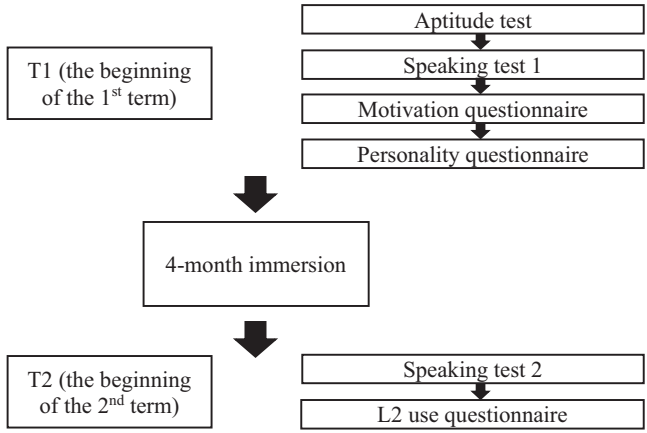

Figure 1 Summary of research time framework. TI = first session; T2 = second session.

we gave the participants 1 minute of planning time; they then proceeded to describe the story within a limit of 2 minutes. To avoid task familiarity effects, the participants described two different sets of pictures (i.e., Cartoons A and B) at T1 and T2, respectively. We counterbalanced the order across the participants (A → B, B → A). The comparison analyses for participants' performance in describing Cartoons A and B suggested that the two sets of pictures were comparable in difficulty level. Appendix S1 in the Supporting Information online provides the materials and results of these analyses. These and other materials for this study are deposited in Tools for Second Language Speech Research and Teaching (http://sla-speech-tools.com; Mora-Plaza et al., 2022) and in the IRIS database (Sun et al., 2023a-c).

We recorded all speech samples with a Roland-05 audio recorder set at 44.1 kHz sampling rate and 16-bit quantization and with a unidirectional condenser microphone. We excised the first 30 seconds of each speech sample (45–83 words) and stored this as a single wav file. We prepared a total of 100 speech samples (50 participants × 2 testing points, T1/T2) and assessed them for the most extensively researched global constructs of L2 speech—comprehensibility and accentedness (Derwing & Munro, 2013).

Comprehensibility and Accentedness Ratings

We recruited three advanced-level L2 English users (L1 Chinese) as raters (3 females; M_{age} = 27.3 years). All held master's degrees and were doctoral

students in the applied linguistics program at Birkbeck, University of London. They had lived in the UK for more than 2.5 years and reported extensive English teaching experiences (around 5 years) with Chinese learners of English. Unlike most L2 speech assessment studies where raters have been native speakers, this study recruited nonnative raters who were highly proficient in the target language, who were familiar with the target L2 accent, and who had extensive metalinguistic awareness (for a similar decision, see Derwing & Munro, 2013). Their backgrounds allowed them to better identify the nuance of L2 accent influenced by L1 Chinese, which could be useful in rating accentedness. Crowther et al. (2016) found no difference in the ratings of comprehensibility and accentedness given by native and nonnative raters with profiles similar as those of the raters in our study (i.e., advanced L2 proficiency, length of residence of about 2 years; for a meta-analysis of rater effects in L2 speech judgements, see Saito, 2021). We believed that using nonnative raters was an advantage as did Hayes-Harb et al. (2008).

The rating sessions took place individually in a quiet room. First, raters received brief instructions about the constructs of comprehensibility and accentedness (see Appendix S2 in the Supporting Information online for training scripts). We defined accentedness as how much speakers' speech is influenced by their native language and/or is colored by other nonnative features (Saito et al., 2017), as a wide range of L1 English accents can be found in London, UK (e.g., standard/estuary/cockney British accent, American accent, Indian accent). Subsequently, raters practiced the procedure with five speech samples at various levels (not included in the main dataset) where they listened to each of the samples and chose a score for comprehensibility and accentedness respectively on screen (see Appendix S2 in the Supporting Information online for onscreen labels). The raters heard each sample only once and rated the samples on a 9-point scale for comprehensibility (1 = *hard to understand*, 9 = *easy to understand*) and accentedness (1 = *heavily accented*, 9 = *no accent*) simultaneously. We encouraged the raters to use the full scale as much as possible. Finally, the raters proceeded to evaluate the 100 speech samples in a randomized order using Praat (Boersma & Weenink, 2012). The entire session took approximately 1 hour, with a 5-min break in the middle.

Cronbach's alpha analyses indicated that the scores given by three raters were highly consistent for comprehensibility ($\alpha = .913$) and accentedness ($\alpha = .930$). Therefore, we calculated the mean of the three raters' scores for each participant's comprehensibility and accentedness at T1 and T2, respectively.

L2 Experience Questionnaire

To obtain a comprehensive estimate of the participants' L2 use between T1 and T2, we administered a questionnaire (Hernández, 2010) adapted from the Language Contact Profile (Freed et al., 2004). Each participant completed the questionnaire with the researcher, reporting the quantity (how many hours they used L2 English per week) and quality of L2 use (interactive vs. noninteractive use; whether the interlocutors with whom they interacted were fluent or not, based on participants' own judgement). We calculated the number of hours that participants spent speaking according to the interlocutor type: (a) interactive L2 use with fluent/L1 speakers, and (b) interactive L2 use with nonfluent speakers. We also calculated (c) noninteractive L2 use as the sum of hours that the participants spent listening, reading, and writing. For noninteractive L2 use, we calculated only out-of-class hours, as the L2 use in class was mostly a mixture of all the three types of activities. Appendix S3 in the Supporting Information online lists the detailed questions. Although the data were self-reported in a retrospective manner, no participant had difficulties recalling the number of hours and responding to the questions, possibly because the participants generally had a fixed schedule during the 4 months.

Aptitude Test

We used the LLAMA test battery (Version 2; Meara, 2005) to measure four domains of participants' aptitude: sound recognition (LLAMA D) for incidental/implicit learning aptitude; associative memory (LLAMA B), phonemic coding (LLAMA E), and grammatical inferencing (LLAMA F) for explicit learning aptitude. An advantage of the LLAMA test (in comparison to the Modern Language Aptitude Test) is its L1 independence, as all verbal materials used in the test are adapted from an indigenous language in North America to which test-takers have likely never been exposed. To minimize participants' intentional learning (being aware of the learning target) in LLAMA D, we administered the test in the following order: LLAMA D → B → E → F. The subtests were automatically scored out of 75 for LLAMA D and 100 for LLAMA B, E, and F.

LLAMA D

The LLAMA D subtest measures the ability to distinguish novel sound sequences from those heard only once in an incidental fashion. To avoid participants' intentional learning, we did not mention the purpose of the test. Instead, we asked the participants to listen to a set of 10 sound strings from the computer to check if they could hear them normally. Then, the participants

proceeded to a recognition test where they listened to 30 sound strings and judged whether they had heard them or not. According to the posttest interviews, no participants reported any intentional memorization of the 10 sound strings at the beginning.

LLAMA B

The LLAMA B subtest assesses the ability to learn new words by associating word strings to unfamiliar object images. First, we asked the participants to click on 20 images and to memorize the displayed word strings within a 2-min time limit. Then, the participants moved to the testing phase where we asked them to match the 20-word strings (appearing in a random order) with the corresponding images.

LLAMA E

The LLAMA E subtest evaluates the ability to associate sound strings with unfamiliar alphabetical symbols. First, we instructed the participants to learn a total of 24 sound–symbol correspondences within 2 minutes (one syllable per symbol). Subsequently, we tested the participants to determine whether they could correctly identify the symbols corresponding to 20 two-syllable sounds out of two choices.

LLAMA F

The LLAMA F subtest measures the ability to deduce the grammatical rules of an unfamiliar language. First, participants had 5 minutes to study 20 pairs of sentences and images and to infer the grammatical rules. We then asked them to choose the grammatically correct sentence from two options for each of the 20 images displayed in a random order.

Motivation Questionnaire

To survey participants' L2 motivation, we administered at T1 an adapted version of the L2 Motivational Self System Questionnaire (Taguchi et al., 2009) that had been validated among Chinese learners in an English L2 context (Q. Li, 2014). The questionnaire consisted of 51 statements, and participants responded to each on a 6-point Likert scale (1 = *strongly disagree*, 6 = *strongly agree*). We used 10 items for ideal L2 self (five items) and ought-to L2 self (five items) in our study. However, following the recent development of the theoretical model of L2MSS (Papi et al., 2019; Teimouri, 2017), we relabeled the original ideal L2 self scale as ideal L2 self/own, given the promotion focus and the standpoint of oneself, and relabeled the original ought-to L2 self

Table 1 Descriptive statistics for L2 comprehensibility and accentedness scores in the first session (T1) and second session (T2)

	T1				T2			
Measure	*M*	*SD*	Range	95% CI	*M*	*SD*	Range	95% CI
Comprehensibility	6.4	1.2	3.0–8.3	[6.1, 6.7]	6.4	1.2	2.0–8.7	[6.1, 6.8]
Accentedness	6.3	1.3	2.7–8.7	[5.9, 6.6]	6.3	1.4	2.3–8.7	[5.9, 6.7]

scales as ideal L2 self/other, given the promotion focus and the standpoint of others. Appendix S4 in the Supporting Information online provides a list of questionnaire items for each scale.

Personality Questionnaire

We surveyed the participants' personality via the short form version of the Multicultural Personality Questionnaire comprising 40 items (van der Zee et al., 2013). We measured five personality traits closely related to the success of multicultural interaction, including cultural empathy, flexibility, social initiative, open-mindedness, and emotional stability. We added a Chinese translation to each item to avoid misunderstandings. We asked the participants to indicate the extent to which each statement was applicable to them based on a 5-point Likert scale (1 = *totally not applicable*, 5 = *completely applicable*; 12 items were reverse-scored). The questionnaire items for all personality traits are listed in Appendix S5 in the Supporting Information online.

Results

Data Preparation

First, we explored the descriptive data to obtain an overall picture of the participants' L2 speech gains and their L2 use and ID profiles.

L2 Speech Gains

As Table 1 shows, the participants' comprehensibility and accentedness rating scores did not change from T1 to T2 at the group level. However, as Figures 2 and 3 depict, the participants' individual changes in both rating scores demonstrated great variability, which allowed us to examine the effects of L2 use and ID variables on L2 speech gains. We calculated gain scores (dependent variable) by deducting T1 scores from T2 scores. To remove one extreme outlier (more than three box-lengths outside the box in the boxplot) in accentedness gain scores, we changed the lowest score from −2.33 to −2.00 (but it remained

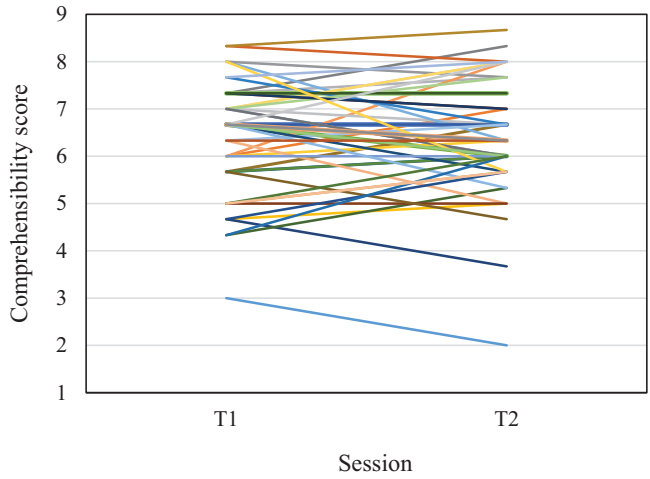

Figure 2 Individual changes in L2 comprehensibility scores over time. T1 = first session; T2 = second session.

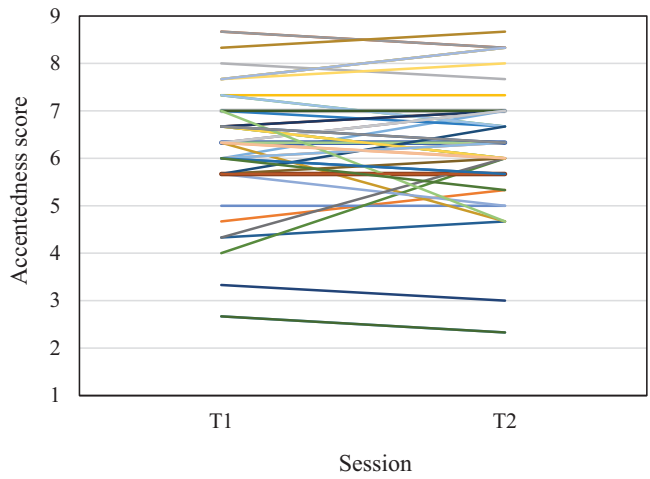

Figure 3 Individual changes in L2 accentedness scores over time. T1 = first session; T2 = second session.

the lowest), as we could not use data transformation due to negative and zero gain scores. To reduce outliers (more than 1.5 box-lengths outside the box in boxplot) in comprehensibility T1 scores (controlled variable), we transformed the data using reflect and square root transformations.

Table 2 Descriptive statistics for the participants' L2 use profiles (hours per week)

L2 use variable	*M*	*SD*	Range	95% CI
Interactive L2 use with fluent speakers	4.5	4.7	0.1–20.0	[3.2, 5.9]
Interactive L2 use with nonfluent speakers	1.3	1.9	0.0–8.0	[0.8, 1.9]
Noninteractive L2 use	36.7	16.9	9.8–83.0	[33.9, 43.5]

Table 3 Descriptive statistics of the participants' cognitive individual difference profiles

Cognitive variable	*M*	*SD*	Range	95% CI
LLAMA D (75 points)	23.0	14.8	0–50	[18.8, 27.2]
LLAMA B (100 points)	50.9	17.9	15–85	[45.8, 56.0]
LLAMA E (100 points)	74.0	27.0	0–100	[66.3, 81.7]
LLAMA F (100 points)	63.8	16.6	30–90	[59.1, 68.5]

Note. LLAMA D = sound recognition; LLAMA B = associative memory; LLMA E = phonemic coding; LLAMA F = grammatical inferencing.

L2 Use

We measured both interactive (speaking) and noninteractive (listening, reading, and writing) L2 use on the basis of the number of hours per week that the participants reported using English during the first 4 months of their study-abroad (T1–T2). The summary in Table 2 shows that learners generally talked in English more with fluent speakers than nonfluent speakers, but they spent much more time using English in a noninteractive manner. We transformed both interactive L2 use measures using the square root transformation to reduce outliers. Although a few outliers remained for interactive L2 use with fluent speakers, they were close to the boundary.

Cognitive Individual Differences

Table 3 lists the participants' aptitude scores measured by four LLAMA subtests. Compared with the guide provided in Meara (2005) and the sample in Bokander and Bylund (2020), the participants obtained relatively high explicit aptitude scores: means of 50.9% to 74% for associative memory (LLAMA B), phonemic coding (LLAMA E), and grammatical inferencing (LLAMA F). In contrast, their incidental/implicit aptitude scores were relatively low, with a mean of 30.7% for sound recognition (LLAMA D). As the total possible scores for the four subtests were different, we converted the participants' aptitude scores into Fisher's *z* scores for the subsequent regression analyses.

Table 4 Descriptive statistics of the participants' sociopsychological individual difference profiles

Sociopsychological variable	*M*	*SD*	Cronbach's α	Range	95% CI
Motivation					
Ideal L2 self/own	4.5	0.7	.64	2.8–5.8	[4.2, 4.7]
Ideal L2 self/other	2.6	1.1	.84	1.0–5.0	[2.3, 2.9]
Personality					
Cultural empathy	4.0	0.5	.82	2.8–4.9	[3.8, 4.1]
Flexibility	2.6	0.7	.82	1.0–4.3	[2.4, 2.8]
Social initiative	3.2	0.7	.85	1.8–4.4	[3.0, 3.4]
Open-mindedness	3.7	0.4	.61	2.8–4.5	[3.5, 3.8]
Emotional stability	3.0	0.7	.84	1.6–4.3	[2.8, 3.2]

Note. Motivation items were rated on a 6-point Likert scale, and personality items were rated on a 5-point Likert scale.

Sociopsychological Individual Differences

Table 4 presents an overview of participants' motivation and personality profiles. We examined the reliability of each motivation and personality scale via the Cronbach's alpha statistic. We deemed their reliability as acceptable; all alphas were greater than .60. For motivation, the participants generally showed stronger ideal L2 self from own standpoint than that from other standpoint. As for personality, the participants appeared to score high on cultural empathy and open-mindedness and relatively low on flexibility.

To examine the validity of sociopsychological ID scales, we submitted items from the two motivation dimensions and five personality dimensions to separate factor analyses. We extracted a total of eight sociopsychological ID factors, consisting of three motivation factors—Ideal L2 Self/Other, Ideal L2 Self/Own-General, and Ideal L2 Self/Own-International, and five personality factors—Cultural Empathy, Emotional Stability, Extraversion, and Orientation to Action (two aspects of the original social initiative scale), and Flexibility. We generated composite scores for each factor (with all loaded items integrated) and used them in the subsequent regression analyses. As the original open-mindedness scale did not remain as a factor, we could not examine its effect on L2 speech gains. See Appendix S6 in the Supporting Information online for more information about the procedure and results of the factor analyses.

Predictors of L2 Speech Gains

Research Question 1 concerned the degree to which L2 experiential and ID variables could predict, collectively and independently, the participants' L2 speech gains over the first 4 months of immersion. To identify potential predictors of L2 gains in comprehensibility and accentedness, we ran a set of hierarchical multiple regression analyses with gain scores as the dependent variable, T1 scores as the first independent variable (to be controlled for), and one of the experiential and ID variables (as calculated in previous sections) as the second independent variable. We tested the assumptions of multiple regression in this section for normal distribution of the errors (via P-P plots of regression standardized residuals) and for multicollinearity (i.e., variance inflation factor < 2).

Comprehensibility Gains

Results showed that in addition to the variance in L2 comprehensibility gain scores explained by T1 scores (10.4%), significant contributions ($p < .05$) could be made to the model by interactive L2 use with fluent speakers, noninteractive L2 use, or grammatical referencing (LLAMA F), accounting for 17.7%, 7.6%, and 7.7% of variance, respectively. We also considered Extraversion as a potential predictor because it showed marginal significance (accounting for 7% of variance, $p = .053$). Thus, we ran a set of hierarchical multiple regression analyses with T1 scores and the four predictors as independent variables. As Table 5 shows, the best fitting model for predicting L2 comprehensibility gains consisted of T1 scores (10.4%), grammatical inferencing (7.7%), Extraversion (8.0%), and interactive L2 use with fluent speakers (12.4%; entered in regression in this order), collectively explaining 38.5% of the variance in gain scores, $R^2 = .385$, $F(4, 45) = 7.037$, $p < .001$, $R^2_{\text{adjusted}} = .330$. The addition of noninteractive L2 use did not significantly improve the variance accounted for by the model regardless of the order in which we entered it into the regression.

Accentedness Gains

Interestingly, accentedness T1 scores alone did not stand as a significant predictor (nor did any other independent variables) of the gain scores. However, when we added LLAMA E (phonemic coding) scores to the regression model as the second independent variable, this accounted for a significant 12.7% of variance in accentedness gains, $R^2 = .127$, $F(2, 47) = 3.415$, $p = .041$, $R^2_{\text{adjusted}} = .090$, with a contribution of 9% from LLAMA E.

Table 5 Multiple regression model predicting L2 speech gains from L2 use and cognitive and sociopsychological individual differences variables

Outcome variable	Predictor variable	*b*	95% CI	*SE*	*t*	*p*	R^2_{change}
Comprehensibility gain scores ($R^2 = .385$)	Constant	−2.85*	[−4.20, −1.51]	0.67	−4.29	< .001	–
	T1 comprehensibility scores	1.30*	[0.64, 1.96]	0.33	3.95	< .001	.104
	LLAMA F (grammatical inferencing)	0.27*	[0.06, 0.48]	0.11	2.57	.014	.077
	Extraversion	0.13	[−0.10, 0.36]	0.11	1.13	.265	.080
	Interactive L2 use with fluent speakers	0.36*	[0.12, 0.60]	0.12	3.01	.004	.124
Accentedness gain scores ($R^2 = .127$)	Constant	1.07*	[0.07, 2.07]	0.50	2.15	.037	–
	T1 accentedness scores	−0.17*	[−0.33, −0.01]	0.08	−2.16	.036	.037
	LLAMA E (phonemic coding)	0.23*	[0.02, 0.44]	0.10	2.20	.033	.090

Note. *Statistic meets alpha = .05. T1 comprehensibility scores were transformed using the reflect function; a positive *b* coefficient indicates a negative correlation. T1 = first session.

According to the field-specific effect size benchmark for global ratings of spontaneous L2 speech, $M = 0.51$, 95% CI [0.30, 0.71], reported by Saito and Plonsky (2019, p. 684), the variance in the L2 gains explained by the experiential variable corresponded to a large effect size, $d = 0.75$, whereas the variance explained by the cognitive and sociopsychological ID variables reached the level of medium effect size: LLAMA F, $d = 0.58$; Extraversion, $d = 0.59$; LLAMA E, $d = 0.63$.[1]

Mediation Effect of Experiential Factors

Research Question 2 concerned whether experiential variables mediate the effect of sociopsychological ID variables on L2 speech gains. On the basis of the result of Research Question 1, we examined whether interactive L2 use with fluent speakers was a mediator in the effect of Extraversion on L2 comprehensibility gains via hierarchical regression analyses. Given that when we entered Extraversion into the regression after interactive L2 use with fluent speakers, it did not significantly explain any extra variance in L2 comprehensibility gain scores (while the L2 use variable explained 18.7% of variance), its effect can be considered as fully mediated by L2 use. The significant result of the Sobel test, $t_{\text{Sobel}} = 2.00$, $SE = 0.07$, $p = .045$, also confirmed the mediation effect. Following MacKinnon et al.'s study (1995), we conducted the test in a regression analysis with the Extraversion predicting the mediator (interactive L2 use with fluent speakers), $b = 0.38$, $SE = 0.13$, $p = .006$, and in a regression analysis with the mediator (and the predictor) predicting the outcome variable (L2 comprehensibility gains), $b = 0.36$, $SE = 0.13$, $p = .007$.

For the other personality variable Emotional Stability and the motivation variables that could encourage L2 use, although none of them significantly predicted L2 comprehensibility gains, it was still worth examining whether and to what degree they predicted interactive L2 use with fluent speakers, which in turn contributed to L2 comprehensibility gains. According to results of the linear regression analyses, we found only Ideal L2 Self/Own-International to be a significant predictor, explaining 12% of variance in the L2 use variable, $R^2 = .120$, $F(1, 48) = 6.517$, $p = .014$, $R^2_{\text{adjusted}} = .101$, with a large effect size, $d = 0.74$. This indicated the possibility of the motivation effect on L2 comprehensibility gains (if any) being mediated by L2 use.

Discussion

The main goal of our study was to explore the degree to which a multidimensional model of L2 experiential, cognitive, and sociopsychological individual differences could explain L2 speech learning in the initial stage of naturalistic

immersion, responding to Moyer's (2014) call for an integrated approach. The results have provided support for such an approach as three types of variables contributed to gains in L2 comprehensibility and accentedness to different degrees and in different ways.

Although the participants made no significant gains at the group level, we observed a great amount of individual variability in both comprehensibility and accentedness gain scores within the group. As we had expected, and in line with previous findings, gains in L2 comprehensibility were explained by not only L2 experience (interactive L2 use with fluent speakers) but also certain cognitive (grammatical inferencing) and sociopsychological (Extraversion) IDs, whereas gains in L2 accentedness were not affected by L2 experience but by a certain pronunciation-related cognitive ID variable (phonemic coding). This study thus highlighted how different dimensions contribute to L2 speech learning—with the experiential variable having a larger effect size than ID variables. Furthermore, the amount of interactive L2 use mediated the effect of sociopsychological IDs (Extraversion, and potentially Ideal L2 Self/Own-International). Though the motivation variable did not predict L2 gains directly, individuals with a clearer image of ideal L2 self in international contexts tended to spend more time speaking with fluent speakers, which in turn contributed to gains in comprehensibility.

Experience

The null result of L2 speech gains at the group level was consistent with findings of Derwing et al.'s (2008) longitudinal study that revealed that L2 comprehensibility of Chinese L2 learners remained unchanged over 8 months of early immersion, whereas the group of Slavic L2 learners, who reported more frequent conversations in L2, showed significant improvement. The small amount of L2 interaction made by the participants ($M = 4.5$ hours per week with fluent speakers; $M = 1.3$ hours with nonfluent speakers) and the short length of residence (4 months) was probably below the threshold for significant group-level gains.

More important, our analyses revealed much individual variability within the group in both L2 speech measures. Both interactive and noninteractive L2 use explained the variance in comprehensibility gains. Compared with noninteractive L2 use and ID variables, interactive L2 use with fluent speakers was the strongest predictor of comprehensibility gains with a large effect size. That is, the participants who had spent more time talking to fluent speakers in L2 (potentially with less time spent using L1) became more comprehensible, whereas those who did not spend sufficient time should even see a decline in

L2 comprehensibility during immersion. Interestingly, the amount of interaction with nonfluent L2 speakers did not matter. This provides further empirical support for Flege's speech learning model (Flege & Bohn, 2021) which conceptualizes the acquisitional value of L2 experience in terms of both quantity and quality (see also Moyer, 2011).

The findings also emphasize that comprehensibility is a more realistic barometer of adult L2 speech learning than accentedness (Munro & Derwing, 2020). In line with Derwing and Munro's (2013) study, accentedness gains were not sensitive to naturalistic L2 experience as much as comprehensibility gains were, given that no experiential variables in this study predicted the variance in accentedness gains. It is likely that accent reduction relies on extensive L2 use and/or targeted instruction (see Saito & Plonsky, 2019, for a review of the effectiveness of L2 pronunciation instruction).

Aptitude

We identified a significant role for explicit aptitude (grammatical inferencing and phonemic coding) in L2 speech gains during early immersion, consistent with findings of previous cross-sectional studies of L2 ultimate attainment in naturalistic (e.g., Granena & Long, 2013) and classroom (e.g., Saito, 2017) settings. The aptitude effect showed a medium effect size according to Saito and Plonsky's (2019) benchmarks. The results led us to propose that adult L2 learners still rely on explicit aptitude to maximize their L2 learning experience, at least at the initial stage of naturalistic immersion, indicating that their learning process is mainly explicit rather than implicit (Suzuki & DeKeyser, 2017). This could be linked to the extensive training of explicit learning through L2 formal instruction in classrooms (10–16 years) before the immersion and the limited L2 experience during the immersion.

More specifically, L2 comprehensibility gains were related to grammatical inferencing ability, and accentedness gains were associated with phonemic coding ability. The findings are compatible with Skehan's (2016) aptitude-acquisition framework. In this view, learners with higher grammatical inferencing ability are assumed to have the capacity to grasp quickly the grammatical patterns in L2 input at morphosyntactic level (Meara, 2005; Saito, 2017) while restructuring their existing systems (Yilmaz, 2013). Therefore, it is possible that such analytical L2 learners could speak a target language with more accurate, complex, and sophisticated grammar and vocabulary that are relevant to the overall comprehensibility of L2 speech (Trofimovich et al., 2022). Phonemic coding ability, however, is linked particularly to the input processing stage of L2 acquisition. Learners better at phonemic coding are

assumed to be able to process the primitive structure of unfamiliar sounds and thus retain them more easily and accurately, which could then contribute to accent reduction. It is noteworthy, however, that regular interaction with fluent speakers showed a larger effect in L2 speech gains than aptitude variables did, indicating that sufficient high-quality L2 experience could be a prerequisite for aptitude to play a role in boosting L2 speech learning.

Motivation

Partially in line with previous studies of the relationship between motivation and L2 speech learning (e.g., Hernández, 2010; Nagle, 2018; Saito et al., 2018), internalized motivation in the current study showed an indirect effect on L2 comprehensibility gains. That is, the participants with a more vivid image of themselves speaking the target language in international contexts (Ideal L2 Self/Own-International) tended to seek more L2 interaction with fluent speakers, which then led to greater gains in comprehensibility. This result is consistent with the idea presented by Papi and Khajavy (2021) that different motivational dimensions may trigger different types of learning behavior, which then affect learning outcome in different ways. To enhance this specific motivational dimension, learners may benefit from training that addresses international posture (i.e., attitudes toward international communities and activities) which has been found as an antecedent of Ideal L2 Self/Own (Yashima, 2002).

Learning behavior likely mediates the influence of motivation on L2 speech learning. However, we found no relationship between Ideal L2 Self/Own-International and L2 comprehensibility gains. We speculate that the ratio of L1 use to L2 use (not measured, unfortunately) may provide some explanation. For example, the participants who spent more hours speaking with fluent speakers in the L2 might have happened to spend even more time speaking L1, which could have confounded the relationship between motivation and L2 comprehensibility gains (see Saito, Macmillan, et al., 2020). Therefore, it is important to track the use of all languages in future studies.

As we expected, the participants reported a low level of Ideal L2 Self/Other ($M = 2.6$ out of 6), probably due to the weaker influence from their closest families and friends in the home country. It may explain why this motivation variable did not significantly predict L2 interaction or L2 comprehensibility gains. Furthermore, given the null effect of interactive L2 use on accent reduction, it is not a surprise that no motivation variables affected accentedness gains. However, future studies may investigate accent-targeted motivation and L2 use to provide more insights.

Personality

To our knowledge, our study was the first attempt in the field of L2 acquisition to conduct a longitudinal examination of the relationship between personality and linguistic gains, hence the exploratory nature of the interpretation of the results. As we had anticipated, we found one interaction-oriented personality trait, Extraversion, that predicted L2 comprehensibility gains indirectly as L2 experience fully mediated its effect. It is not surprising that extraverts (i.e., who were more likely to engage in social interaction) spent more time talking with fluent speakers in the target language, which in turn resulted in more gains in comprehensibility. However, although individuals who were more emotionally stable might have more positive attitudes towards the L2 community (MacIntyre & Charos, 1996) and be more comfortable about L2 interaction (Dewaele & Al-Saraj, 2015; Pyun et al., 2014), these characteristics did not get learners involved in more L2 interaction as Extraversion did. Unfortunately, we could not examine the effect of Open-Mindedness as it was removed during the factor analyses, arguably due to the high scores on average ($M = 3.7$ out of 5) and thus the lack of variance in this dataset. Similarly, the overall high level of Cultural Empathy ($M = 4.0$) could have confounded its effect on L2 accentedness gains that previous studies have found (Ibrahim et al., 2008; Rota & Reiterer, 2009).

Limitations and Future Directions

Our study took a first step in applying the multidimensional approach to L2 speech learning in naturalistic settings with a longitudinal design. Although the results confirmed the importance of an integrated consideration of experiential, cognitive, and sociopsychological dimensions, the amount of variance in L2 comprehensibility gains explained by the three types of variables collectively could be considered as medium (28%, with around 12% explained by L2 experience and 8% by cognitive or sociopsychological IDs), compared with L2 studies in general (Plonsky & Ghanbar, 2018).

One reason could be that we measured all the variables only once, which made it impossible to show the dynamic fluctuation during the project. Thus, future studies should aim at multiple data collection episodes over an extensive period of immersion, especially for less static constructs such as L2 experience and motivation. For example, instead of retrospective reporting, participants' L2 experience could be surveyed more frequently and in depth by using advanced tools such as the Lang-Track-App developed by Arndt et al. (2022). This smart phone application can be adapted to the purpose of data collection, so questions about L2 motivation might also be included. Even for widely

acknowledged stable dimensions such as aptitude and personality traits (Hu & Reiterer, 2009; Skehan, 2002), it might be worth testing the assumptions via qualitative analyses from longitudinal perspectives.

Another direction for future research is to expand the experience/ID framework by incorporating more varied and fine-grained instruments. Following previous studies on experience effects (Derwing & Munro, 2013; Flege & Liu, 2001), we measured experience in quantity (hours per week) and sorted it out according to the mode of activities (interactive vs. noninteractive) and types of interlocutors (fluent vs. nonfluent). However, recent studies have suggested that the context of L2 use (professional/academic vs. social vs. home) and the L1–L2 use ratio may also influence L2 speech learning (Moyer, 2011; Saito, Macmillan, et al., 2020).

Similarly, although we examined ID variables via state-of-art instruments, after we had completed the data collection for this study, questions have been raised about some of those measures. In a recently published validation study on the aptitude test of LLAMA (Version 2), it was reported that almost all subtests failed to reach good validity except LLAMA B (Bokander & Bylund, 2020). A revised version of the tests, LLAMA (Version 3), is a work in progress (Meara & Rogers, 2019). Thus, future studies will need to use more valid aptitude measures (e.g., domain-general auditory processing measures; Saito, Sun, & Tierney, 2020) to examine the multidimensional model. The L2MSS model has also been criticized for its simplicity and refined by Teimouri (2017) and Papi et al. (2019) to include the division of personal aspect (own) and social aspect (others) and with clear regulatory distinctions (promotion for ideal L2 self vs. prevention for ought-to L2 self). Due to the limitation in our dataset, we could not investigate the effect of ought-to L2 self, as the items were relabeled as ideal L2 self/other according to the revised model. Furthermore, more context-specific motivational scales could be integrated to target L2 speech learning, such as motivation related specifically to L2 comprehensibility and accentedness development (Saito et al., 2018).

For future studies, the cognitive variables could be further expanded to feature cognitive abilities typical of advanced L2 learners in naturalistic settings, such as auditory processing abilities (Sun et al., 2021), phonological short-term memory (Foster et al., 2014), working memory (Linck et al., 2013), and implicit learning abilities (Granena, 2013). The sociopsychological orientations of L2 learners could be expanded to include constructs which may interact with motivation or personality, such as emotion (Dewaele & MacIntyre, 2014; Dewaele et al., 2022, for anxiety and enjoyment), willingness to

communicate (MacIntyre, 2007), international posture (Yashima et al., 2017), and ethnic identity (e.g., Pavlenko & Blackledge, 2004).

Finally, our study was one of the first attempts to take an integrated approach toward disentangling the complex roles of cognitive (aptitude), sociopsychological (motivation, personality), and experiential individual differences (quantity and quality of language use) in L2 speech learning within a single study. We call for future studies which will not only replicate the findings with other groups of L2 learners with different biographical profiles (e.g., age, emotion, L1–L2 pairings) but also extend the framework to classroom settings with a view of more directly relevant pedagogical implications (cf. Suzukida & Saito, 2022). With such studies, researchers will test our hypotheses (a) that comprehensibility is a more dynamic construct of L2 speech that constantly develops as per L2 learners' motivation, emotion, and language use (Trofimovich et al., 2020) and (b) that accentendess is a more invariable construct of L2 speech that is resistant to change owing to its close link with aptitude (Saito et al., 2019).

Conclusion

Our study examined how L2 comprehensibility and accentedness of 50 adult Chinese learners of English changed during a short period of initial immersion in relation to their individual differences in aptitude, motivation, personality, and L2 experience. The participants did not make significant gains as a group, but there was great individual variability. By and large, the results echoed the dominant view in L2 speech literature that accessing high-quality interactive experience (speaking with fluent speakers) is fundamental to L2 speech acquisition (e.g., Flege & Bohn, 2021). Furthermore, the results supported the multidimensional model that Moyer (2014) proposed, that is, that variance in adult L2 speech learning can be explained not only by experiential variables but also by learner-internal cognitive and sociopsychological ID variables to different degrees and in different ways. Here, we have provided a tentative conclusion regarding the mechanism underlying different dimensions of adult L2 speech learning. Whereas improvement in communication-oriented dimension (i.e., comprehensibility) is primarily driven by experience, such experience–acquisition links could be particularly strong among L2 learners with certain cognitive ID profiles—those with stronger lexicogrammar-related aptitude tend to make more gains. Sociopsychological IDs influence acquisition indirectly via the mediation of experience—those who are more extraverted or have a clearer image of their ideal L2 self in international contexts tend to have more high-quality L2 interaction. Comparatively, the

extent to which L2 learners can approximate a nativelike accent tends to be less subject to experience effects but to be affected more by certain ID dimensions (e.g., pronunciation-related aptitude).

Final revised version accepted 3 January 2023

Open Research Badges

This article has earned an Open Materials badge for making publicly available the components of the research methods needed to reproduce the reported procedure. All materials that the authors have used and have the right to share are available at http://sla-speech-tools.com and http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

Note

1 We converted the effect size measure from R^2_{change} to Cohen's d using the formula $r = d/\sqrt{(d^2+4)}$ proposed by Ruscio (2008).

References

Al-Hoorie, A. H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*(4), 721–754. https://doi.org/10.14746/ssllt.2018.8.4.2

Arndt, H. L., Granfeldt, J., & Gullberg, M. (2022). The Lang-Track-App: Open-source tools for implementing the experience sampling method in second language acquisition research. Advance online publication. *Language Learning*. https://doi.org/10.1111/lang.12555

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer*. http://www.praat.org

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude Tests. *Language Learning*, *70*(1), 11–47. https://doi.org/10.1111/lang.12368

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, *55*, 145–157. https://doi.org/10.1016/j.system.2015.10.006

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. Psychological Corporation.

Chen, X., He, J., Swanson, E., Cai, Z., & Fan, X. (2022). Big five personality traits and second language learning: A meta-analysis of 40 years' research. *Educational Psychology Review*, *34*, 851–887. https://doi.org/10.1007/s10648-021-09641-6

Costa, P., & McCrae, R. (1989). *The NEO-PI/NEO-FFI manual supplement*. Psychological Assessment Resources.

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation*, *2*(2), 160–182. https://doi.org/10.1075/jslp.2.2.02cro

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163–185. https://doi.org/10.1111/lang.12000

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380. https://doi.org/10.1093/applin/amm041

Dewaele, J.-M., & Al-Saraj, T. M. (2015). Foreign language classroom anxiety of Arab learners of English: The effect of personality, linguistic and sociobiographical variables. *Studies in Second Language Learning and Teaching*, *5*(2), 205–228. https://doi.org/10.14746/ssllt.2015.5.2.2

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, *4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., Saito, K., & Halimi, F. (2022). How foreign language enjoyment acts as a buoy for sagging motivation: A longitudinal investigation. *Applied Linguistics*. Advance online publication. https://doi.org/10.1093/applin/amac033

Dormann, C., & Guthier, C. (2019). Longitudinal data collection. In P. Brough (Ed.), *Advanced research methods for applied psychology: Design, analysis and reporting* (pp. 146–157). Routledge. https://doi.org/10.4324/9781315517971-15

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum. https://doi.org/10.4324/9781410613349

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters. https://doi.org/10.21832/9781847691293-003

EIKEN Foundation of Japan. (2016). *EIKEN Pre-1 Level: Complete questions collection*. Oubunsha.

Flege, J., & Bohn, O-S. (2021). The revised speech learning model. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527–552. https://doi.org/10.1017/s0272263101004041

Foster, P., Bolibaugh, C., & Kotula, A. (2014). Knowledge of nativelike selections in a L2: The influence of exposure, memory, age of onset, and motivation in foreign

language and immersion settings. *Studies in Second Language Acquisition*, *36*(1), 101–132. https://doi.org/10.1017/s0272263113000624

Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, *26*(2), 349–356. https://doi.org/10.1017/s027226310426209x

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (pp. 141–165). Sage.

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*(4), 665–703. https://doi.org/10.1111/lang.12018

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and HI-LAB. *Studies in Second Language Acquisition*, *41*(2), 313–336. https://doi.org/10.1017/s0272263118000256

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Guiora, A., Brannon, R., & Dull, C. (1972). Empathy and second language learning. *Language Learning*, *22*(1), 111–130. https://doi.org/10.1111/j.1467-1770.1972.tb00077.x

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36*(4), 664–679. https://doi.org/10.1016/j.wocn.2008.04.002

Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal*, *94*(4), 600–617. https://doi.org/10.1111/j.1540-4781.2010.01053.x

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319–340. https://doi.org/10.1037/0033-295x.94.3.319

Hu, X., & Reiterer, S. (2009). Personality and pronunciation talent in second language acquisition. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and brain activity* (pp. 97–130). Mouton de Gruyter. https://doi.org/10.1515/9783110215496.97

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning*, *71*(3), 626–668. https://doi.org/10.1111/lang.12451

Ibrahim, R., Eviatar, Z., & Leikin, M. (2008). Speaking Hebrew with an accent: Empathic capacity or other nonpersonal factors. *International Journal of Bilingualism*, *12*(3), 195–207. https://doi.org/10.1177/1367006908098566

Kermad, A. (2018). *The effect of individual differences on English learners' oral performance.* (Unpublished doctoral dissertation). Northern Arizona University.

Li, Q. (2014). Differences in the motivation of Chinese learners of English in a foreign and second language context. *System*, *42*, 451–461. https://doi.org/10.1016/j.system.2014.01.011

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634–654. https://doi.org/10.1111/j.1540-4781.2013.12030.x

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*(3), 530–566. https://doi.org/10.1111/lang.12011

MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, *91*(4), 564–576. https://doi.org/10.1111/j.1540-4781.2007.00623.x

MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, *15*(1), 3–26. https://doi.org/10.1177/0261927x960151001

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*(1), 41–62. https://doi.org/10.1207/s15327906mbr3001_3

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954–969. https://doi.org/10.1037/0003-066x.41.9.954

Meara, P. M. (2005). *LLAMA language aptitude tests*. Lognostics.

Meara, P. M., & Rogers, V. E. (2019). *The LLAMA Tests v3*. Lognostics.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J-M., & Tierney, A. (2022). Tools for second language speech research and teaching. http://sla-speech-tools.com

Moyer, A. (2011). An investigation of experience in L2 phonology: Does quality matter more than quantity? *Canadian Modern Language Review*, *67*(2), 191–216. https://doi.org/10.3138/cmlr.67.2.191

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge University Press. https://doi.org/10.1017/cbo9780511794407

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*(4), 418–440. https://doi.org/10.1093/applin/amu012

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, *6*(3), 283–309. https://doi.org/10.1075/jslp.20038.mun

Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development.

The Modern Language Journal, *102*(1), 199–217. https://doi.org/10.1111/modl.12461

Ortega, L. (2009). *Understanding second language acquisition*. Routledge. https://doi.org/10.4324/9780203777282

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45. https://doi.org/10.1017/s0267190505000024

Ożańska-Ponikwia, K., & Dewaele, J.-M. (2012). Personality and L2 use: The advantage of being openminded and self-confident in an immigration context. *EUROSLA Yearbook*, *12*(1), 112–134. https://doi.org/10.1075/eurosla.12.07oza

Papi, M., Bondarenko, A. V., Mansouri, S., Feng, L., & Jiang, C. (2019). Rethinking L2 motivation research: The 2 × 2 model of L2 self-guides. *Studies in Second Language Acquisition*, *41*(2), 337–361. https://doi.org/10.1017/s0272263118000153

Papi, M., & Khajavy, G. H. (2021). Motivational mechanisms underlying second language achievement: A regulatory focus perspective. *Language Learning*, *71*(2), 537–572. https://doi.org/10.1111/lang.12443

Pavlenko, A., & Blackledge, A. (2004). *Negotiation of identities in multilingual contexts*. Multilingual Matters. https://doi.org/10.21832/9781853596483

Pervin, L. A., & Cervone, D. (2010). *Personality: Theory and research*. John Wiley.

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, *102*(4), 713–731. https://doi.org/10.1111/modl.12509

Pyun, D. O., Kim, J. S., Cho, H. Y., & Lee, J. H. (2014). Impact of affective variables on Korean as a foreign language learners' oral achievement. *System*, *47*, 53–63. https://doi.org/10.1016/j.system.2014.09.017

Rota, G., & Reiterer, S. (2009). Cognitive aspects of pronunciation talent. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and brain activity* (pp. 67–96). Mouton de Gruyter. https://doi.org/10.1515/9783110215496.67

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*(1), 19–30. https://doi.org/10.1037/1082-989x.13.1.19

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*(3), 563–595. https://doi.org/10.1111/lang.12120

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, *67*(3), 665–693. https://doi.org/10.1111/lang.12244

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*(3), 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., Dewaele, J.-M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*(3), 709–743. https://doi.org/10.1111/lang.12297

Saito, K., Macmillan, K., Mai, T., Suzukida, Y., Sun, H., Magne, V., Ilkan, M., & Murakami, A. (2020). Developing, analyzing and sharing multivariate datasets: Individual differences in L2 learning revisited. *Annual Review of Applied Linguistics*, *40*, 9–25. https://doi.org/10.1017/s0267190520000045

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Saito, K., Sun, H., & Tierney, A. (2020). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, *41*(5), 1083–1112. https://doi.org/10.1017/s0142716420000491

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*(1), 201–225. https://doi.org/10.1017/s0272263117000432

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462. https://doi.org/10.1093/applin/amv047

Skehan, P. (2002) Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed learning* (pp. 69–93). John Benjamins. https://doi.org/10.1075/lllt.2.06ske

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 17–40). John Benjamins. https://doi.org/10.1075/bpa.3.02ske

Sun, H., Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, *43*(3), 551–573. https://doi.org/10.1017/s0272263120000649

Sun, H., Saito, K. & Dewaele, J. (2023a). *L2 English experience questionnaire from "Cognitive and socio-psychological individual differences, experience, and naturalistic second language speech learning: A longitudinal study"* [Questionnaire]. IRIS Database, University of York, UK. https://doi.org/10.48316/XFtnc-7l4lG

Sun, H., Saito, K. & Dewaele, J. (2023b). *Multicultural personality questionnaire from "Cognitive and socio-psychological individual differences, experience, and naturalistic second language speech learning: A longitudinal study"*

[Questionnaire]. IRIS Database, University of York, UK. https://doi.org/10.48316/xRWJN-3lXWM

Sun, H., Saito, K. & Dewaele, J. (2023c). *Picture narrative tasks from "Cognitive and socio-psychological individual differences, experience, and naturalistic second language speech learning: A longitudinal study"* [Language task]. IRIS Database, University of York, UK. https://doi.org/10.48316/vKMNE-ukqPf

Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition*, *43*(3), 663–676. https://doi.org/10.1017/S0272263120000704

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747–790. https://doi.org/10.1111/lang.12241

Suzukida, Y., & Saito, K. (2022). Disentangling experiential, cognitive, and sociopsychological individual differences in second language speech learning: Cross-sectional and longitudinal investigations. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66–97). Multilingual Matters. https://doi.org/10.21832/9781847691293-005

Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, *39*(4), 681–709. https://doi.org/10.1017/s0272263116000243

Trofimovich, P., Isaacs, T., Kennedy, S., & Tsunemoto, A. (2022). Speech comprehensibility. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 174–187). Routledge. https://doi.org/10.4324/9781003022497-16

Trofimovich, P., Nagle, C. L., O'Brien, M. G., Kennedy, S., Reid, K. T., & Strachan, L. (2020). Second language comprehensibility as a dynamic construct. *Journal of Second Language Pronunciation*, *6*(3), 430–457. https://doi.org/10.1075/jslp.20003.tro

van der Zee, K. I., & van Oudenhoven, J. P. (2000). The Multicultural Personality Questionnaire: A multidimensional instrument of multicultural effectiveness. *European Journal of Personality*, *14*(4), 291–309. https://doi.org/10.1002/1099-0984(200007/08)14:4%3C291::AID-PER377%3E3.0.CO;2-6

van der Zee, K. I., van Oudenhoven, J. P., Ponterotto, J. G., & Fietzer, A. W. (2013). Multicultural Personality Questionnaire: Development of a short form. *Journal of Personality Assessment*, *95*(1), 118–124. https://doi.org/10.1080/00223891.2012.718302

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal*, *86*(1), 54–66. https://doi.org/10.1111/1540-4781.00136

Yashima, T., Nishida, R., & Mizumoto, A. (2017). Influence of learner beliefs and gender on the motivating power of L2 selves. *The Modern Language Journal*, *101*(4), 691–711. https://doi.org/10.1111/modl.12430

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, *34*(3), 344–368. https://doi.org/10.1093/applin/ams044

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary
Appendix S1. Speaking Tasks.
Appendix S2. Training Scripts and Onscreen Labels for L2 Speech Measures.
Appendix S3. L2 Experience Questionnaire.
Appendix S4. Motivation Questionnaire.
Appendix S5. Personality Questionnaire.
Appendix S6. Procedure and Results of Factor Analyses on Sociopsychological Individual Differences.